conclusion that this transfer would be driven by competition is the product of its mistaken economic analysis.

Ch. 5

Investment under Uncertainty

Avinash K. Dixit and Robert S. Pindyck

Princeton University Press
Princeton, New Jersey

must decide when to invest in a single project. The cost of the investment, I, is known and fixed, but the value of the project, V, follows a geometric Brownian motion. The simple net present value rule is to invest as long as $V > I$, but as McDonald and Siegel demonstrated, this is incorrect. Because *future* values of V are unknown, there is an opportunity cost to investing today. Hence the optimal investment rule is to invest when V is at least as large as a critical value V^* that exceeds I. As we will see, for reasonable parameter values, this critical value may be *two or three times* as large as I. Hence the simple NPV rule is not just wrong; it is often *very* wrong.

After describing the basic model in more detail, we will show how the optimal investment rule (that is, the critical value V^*) can be found by dynamic programming. An issue that arises, however, is the choice of discount rate. If capital markets are "complete" (in a sense that will be made clear), the investment problem can be viewed as a problem in option pricing, and solved using the techniques of contingent claims analysis. We will re-solve the optimal investment problem in this way, and then examine the characteristics of the firm's option to invest and its dependence on key parameters. Finally, we will extend the model by considering alternative stochastic processes for the value of the project, V. In particular, we will find and characterize the optimal investment rules that apply when V follows a mean-reverting process, and then when it follows a mixed Brownian motion/Poisson jump process.

1 The Basic Model

Our starting point is a model first developed by McDonald and Siegel (1986). They considered the following problem: At what point is it optimal to pay a sunk cost I in return for a project whose value is V, given that V evolves according to the following geometric Brownian motion:

$$dV = \alpha V\, dt + \sigma V\, dz, \tag{1}$$

where dz is the increment of a Wiener process. Equation (1) implies that the current value of the project is known, but future values are lognormally distributed with a variance that grows linearly with the time horizon; the exact formulas are in Section 3(a) of Chapter 3. Thus although information arrives over time (the firm observes V changing), the future value of the project is *always* uncertain.

Equation (1) is clearly an abstraction from most real projects. For example, suppose the project is a widget factory with some capacity. If variable

costs are positive and managers have the option to shut down the factory temporarily when the price of output is below variable cost, and/or the option to abandon the project completely, V will not follow a geometric Brownian motion even if the price of widgets does. (We will develop models in which the output price follows a geometric Brownian motion and the project can be temporarily shut down and/or abandoned in Chapters 6 and 7.) If variable cost is positive and managers do *not* have the option to shut down (perhaps because of regulatory constraints), V can become negative, which is again in conflict with the assumption of lognormality. In addition, one might believe that a competitive product market will prevent the price from wandering too far from long-run industry-wide marginal cost, or that stochastic changes in price are likely to be infrequent but large, so that V should follow a mean-reverting or jump process. For the time being we ignore these possibilities in order to provide the simplest introduction to the basic ideas and techniques. We allow exogenously specified mean reversion in Section 5(a) of this chapter, and consider industry equilibrium in Chapters 8 and 9.

Note that the firm's investment opportunity is equivalent to a perpetual call option—the right but not the obligation to buy a share of stock at a prespecified price. Therefore the decision to invest is equivalent to deciding when to exercise such an option. Thus, the investment decision can be viewed as a problem of option valuation (as we saw in the simple examples presented in Chapter 2).[1] Alternatively, it can be viewed as a problem in dynamic programming. We will derive the optimal investment rule in two ways, first using dynamic programming, and then using option pricing (contingent claims) methods. This will allow us to compare these two approaches and the assumptions that each requires. We will then examine the characteristics of the solution.

In what follows, we will denote the value of the investment opportunity (that is, the value of the option to invest) by $F(V)$. We want a rule that maximizes this value. Since the payoff from investing at time t is $V_t - I$, we want to maximize its expected present value:

$$F(V) = \max \mathcal{E}[(V_T - I)e^{-\rho T}], \tag{2}$$

where $\mathcal{E}$ denotes the expectation, T is the (unknown) future time that the investment is made, ρ is a discount rate, and the maximization is subject to

[1] The investment opportunity is analogous to a perpetual call option on a dividend-paying stock. (The payout stream from the completed project is equivalent to the dividend on the stock.) A solution to this option valuation and exercise problem was first found by Samuelson (1965).

a critical value V^* such that it is optimal to invest once $V \geq V^*$. As we will see, a higher value of σ will result in a higher V^*, that is, a greater value to waiting. It is important to keep in mind, however, that in general both growth ($\alpha > 0$) *and* uncertainty ($\sigma > 0$) can create a value to waiting and thereby affect investment timing.

In the next two sections, we will solve this investment problem in two ways, following the techniques described in Chapter 4. First, we will use dynamic programming, and then we will solve the same problem over again using contingent claims methods. This will enable us to carefully compare these two approaches.

2 Solution by Dynamic Programming

In the terminology of Chapter 4, we have an optimal stopping problem in continuous time. Because the investment opportunity, $F(V)$, yields no cash flows up to the time T that the investment is undertaken, the only return from holding it is its capital appreciation. Hence, as we saw in Chapter 4, in the continuation region (values of V for which it is not optimal to invest) the Bellman equation is

$$\rho F \, dt = \mathcal{E}(dF). \tag{7}$$

Equation (7) just says that over a time interval dt, the total expected return on the investment opportunity, $\rho F \, dt$, is equal to its expected rate of capital appreciation.

We expand dF using Ito's Lemma, and we use primes to denote derivatives, for example, $F' = dF/dV$, $F'' = d^2F/dV^2$, etc. Then

$$dF = F'(V)\,dV + \tfrac{1}{2} F''(V)\,(dV)^2.$$

Substituting equation (1) for dV into this expression and noting that $\mathcal{E}(dz) = 0$ gives

$$\mathcal{E}[dF] = \alpha V F'(V)\,dt + \tfrac{1}{2}\sigma^2 V^2 F''(V)\,dt.$$

Hence the Bellman equation becomes (after dividing through by dt):

$$\tfrac{1}{2}\sigma^2 V^2 F''(V) + \alpha V F'(V) - \rho F = 0. \tag{8}$$

It will be easier to analyze the solution and to compare it to that obtained using contingent claims analysis if we make the substitution $\alpha = \rho - \delta$. To ensure existence of an optimum (for reasons already explained in connection with

the deterministic case), we assume that $\alpha < \rho$, or $\delta > 0$. With this notation, the Bellman equation becomes the following differential equation that must be satisfied by $F(V)$:

$$\tfrac{1}{2}\sigma^2 V^2 F''(V) + (\rho - \delta) V F'(V) - \rho F = 0. \tag{9}$$

In addition, $F(V)$ must satisfy the following boundary conditions:

$$F(0) = 0, \tag{10}$$

$$F(V^*) = V^* - I, \tag{11}$$

$$F'(V^*) = 1. \tag{12}$$

Condition (10) arises from the observation that if V goes to zero, it will stay at zero [this is an implication of the stochastic process (1) for V]. Therefore the option to invest will be of no value when $V = 0$. The other two conditions come from consideration of optimal investment. V^* is the price at which it is optimal to invest, or in the language of Chapter 4, the free boundary of the continuation region. Then (11) is the value-matching condition; it just says that upon investing, the firm receives a net payoff $V^* - I$. Finally, condition (12) is the "smooth-pasting" condition, discussed in Chapter 4 and its Appendix C. If $F(V)$ were not continuous and smooth at the critical exercise point V^*, one could do better by exercising at a different point.

Note that equation (9) is a second-order differential equation, but there are three boundary conditions that must be satisfied. The reason is that although the position of the first boundary ($V = 0$) is known, the position of the second boundary is not. In other words, the "free boundary" V^* must be determined as part of the solution. That needs the third condition.

Equation (11) has another useful interpretation. Write it as $V^* - F(V^*) = I$. When the firm invests, it gets the project valued V, but gives up the *opportunity* or option to invest, which is valued at $F(V)$. Thus its gain, net of the opportunity cost, is $V - F(V)$. The critical value V^* is where this net gain equals the direct or tangible cost of investment, I. Equivalently, we could write the equation as $V^* = I + F(V^*)$, setting the value of the project equal to the full cost (direct cost plus opportunity cost) of making the investment. We will discuss this point in more detail later.

To find $F(V)$, we must solve equation (9) subject to the boundary conditions (10)–(12). In this case a solution is easy to find; we can guess a functional form, and determine by substitution if it works. We first state the solution and derive some of its properties, and then discuss it in more detail.

the quadratic expression totally:

$$\frac{\partial Q}{\partial \beta}\frac{\partial \beta_1}{\partial \sigma} + \frac{\partial Q}{\partial \sigma} = 0,$$

where all derivatives are evaluated at β_1. Figure 5.2 shows that $\partial Q/\partial \beta > 0$ at β_1. Also

$$\partial Q/\partial \sigma = \sigma\, \beta(\beta - 1) > 0$$

at $\beta_1 > 1$. Therefore $\partial \beta_1/\partial \sigma < 0$. In other words, as σ increases, β_1 decreases, and therefore $\beta_1/(\beta_1 - 1)$ increases. The greater is the amount of uncertainty over future values of V, the larger is the wedge between V^* and I, that is, the larger is the excess return the firm will demand before it is willing to make the irreversible investment.

Readers can likewise verify two other properties of this quadratic. First, β_1 increases as δ increases, so a higher δ means a lower wedge $\beta_1/(\beta_1 - 1)$. Second, β_1 decreases as ρ increases, so a higher ρ implies a larger wedge. We will discuss these results in greater detail and offer some numerical values in Section 4 of this chapter.

Some limiting results concerning β_1 are also informative. We merely state them; they are easily verified using the algebraic formula. First, as $\sigma \to \infty$, we have $\beta_1 \to 1$ and $V^* \to \infty$, that is, the firm never invests if σ is infinite. Next consider what happens as $\sigma \to 0$. We have

If $\alpha > 0$, then $\beta_1 \to \rho/(\rho - \delta)$ and $V^* \to (\rho/\delta)\, I$.

If $\alpha \leq 0$, then $\beta_1 \to \infty$ and $V^* \to I$.

These results conform to those of the deterministic case that we examined earlier.

2.B Relationship to Neoclassical Investment Theory

To push this analysis a bit further, suppose that the project itself is an infinitely lived factory that produces a profit flow, π_t, that follows the process

$$d\pi = \alpha\, \pi\, dt + \sigma\, \pi\, dz.$$

Hence V is given by

$$V_t = \mathcal{E} \int_t^{\infty} \pi_s\, e^{-\rho(s-t)}\, ds = \frac{\pi_t}{\rho - \alpha},$$

and dV is given by equation (1). The usual Marshallian rule is to invest as long as $V_t \geq I$, or $\pi_t \geq (\rho - \alpha) I$. However, equation (14) tells us that instead the firm should invest when

$$\pi_t \geq \pi^* = \frac{\beta}{\beta - 1}\,(\rho - \alpha)\, I > (\rho - \alpha)\, I. \tag{17}$$

Another way to look at this is in terms of the Jorgensonian approach to investment.[3] From the quadratic equation (16) satisfied by β_1, we have

$$\frac{\beta_1}{\beta_1 - 1}\,(\rho - \alpha) = \rho + \tfrac{1}{2}\sigma^2 \beta_1.$$

Thus the critical profit level π^* can be written as

$$\pi^* = (\rho + \tfrac{1}{2}\sigma^2 \beta_1)\, I > \rho\, I. \tag{18}$$

Since we have assumed zero depreciation, ρI is the Jorgensonian user cost of capital; the Jorgensonian rule is to invest when $\pi_t = \rho I$. Equation (18) says that when future profits are uncertain, the threshold π^* must exceed this user cost of capital.

In the absence of uncertainty, the Jorgensonian investment rule has the firm investing when $\pi_t = \rho I$, *not* when $\pi_t = (\rho - \alpha) I$. As we saw before, this can be viewed as an optimal timing rule. Once again, the firm must choose T to maximize

$$\max_T \left(\frac{\pi_0 e^{\alpha T}}{\rho - \alpha} - I \right) e^{-\rho T} = \frac{\pi_0 e^{-(\rho-\alpha)T}}{\rho - \alpha} - I e^{-\rho T}. \tag{19}$$

The solution is to invest at a time T when

$$\pi_T = \pi_0 e^{\alpha T} = \rho\, I \tag{20}$$

(The reader can verify that $\alpha > 0$ is the second-order condition for this maximization.) Therefore the firm should wait to invest even if there is no uncertainty, because waiting allows the postponement (and thus discounting) of the payment I.[4] As equation (18) shows, with uncertainty there is an additional $\frac{1}{2}\sigma^2 \beta_1$ term, so that the firm must wait even longer before investing. This additional term can be thought of as a correction to the neoclassical investment model.

[3]Jorgenson (1963) showed that absent uncertainty, the firm should invest when the marginal profit from an extra unit of capital equals the user cost of capital. Our thanks to Giuseppe Bertola for suggesting this viewpoint in this context.

[4]To our knowledge, this point was first noted by Marglin (1963, Chapter 2).

goods to the extent that prices are correlated with the values of shares or portfolios. However, there may be cases in which this assumption will not hold; an example might be a project to develop a new product that is unrelated to any existing ones, or an R&D venture, the results of which may be hard to predict.

We will assume in this section that spanning holds, that is, that in principle the uncertainty over future values of V can be replicated by existing assets. With this assumption, we can determine the investment rule that maximizes the firm's market value without making any assumptions about risk preferences or discount rates. Also, the use of contingent claims analysis will make it easier to interpret certain properties of the solution. Of course, if spanning does not hold, dynamic programming can still be used to maximize the present value of the firm's expected flow of profits, subject to an arbitrary discount rate. See the discussion in Chapter 4, Section 3 for more on the relationship between the two approaches.

We follow the theory of contingent claims valuation outlined in Chapter 4, Section 2, but repeat some details for reinforcement and clarity. Let x be the price of an asset or dynamic portfolio of assets perfectly correlated with V, and denote by ρ_{xm} the correlation of x with the market portfolio. Since x is perfectly correlated with V, $\rho_{xm} = \rho_{Vm}$. We will assume that this asset or portfolio pays no dividends, so its entire return is from capital gains. Then x evolves according to

$$dx = \mu\, x\, dt + \sigma\, x\, dz, \tag{22}$$

where μ, the drift rate, is the expected rate of return from holding this asset or portfolio of assets. According to the Capital Asset Pricing Model (CAPM), μ should reflect the asset's systematic (nondiversifiable) risk. As explained in Chapter 4, μ will be given by

$$\mu = r + \phi\, \rho_{xm}\, \sigma,$$

where r is the risk-free interest rate, and ϕ is the market price of risk.[6] Thus, μ is the risk-adjusted expected rate of return that investors would require if they are to own the project. We will assume that α, the expected percentage rate of change of V, is less than this risk-adjusted return μ. (As we will see, the firm would never invest if this were not the case. No matter what the current

[6] That is, $\phi = (r_m - r)/\sigma_m$, where r_m is the expected return on the market, and σ_m is the standard deviation of that return. If we take the New York Stock Exchange Index as the market, $r_m - r \approx 0.08$ and $\sigma_m \approx 0.2$, so $\phi \approx 0.4$. For a more detailed discussion of the Capital Asset Pricing Model, see Brealey and Myers (1991) or Duffie (1992).

level of V, the firm would always be better off waiting and simply holding on to its option to invest.) We will let δ denote the difference between μ and α, that is, $\delta = \mu - \alpha$. Thus we are assuming $\delta > 0$, and this plays the same role as the corresponding assumption in the dynamic programming formulation of Section 2.

The parameter δ plays an important role in this model. We discussed its role as an explicit or implicit dividend in Chapter 4; here we elaborate on those remarks. It will be helpful to draw upon the analogy with a financial call option. If V were the price of a share of common stock, δ would be the dividend rate on the stock. The total expected return on the stock would be $\mu = \delta + \alpha$, that is, the dividend rate plus the expected rate of capital gain. If the dividend rate δ were zero, a call option on the stock would always be held to maturity, and never exercised prematurely. The reason is that the entire return on the stock is captured in its price movements, and hence by the call option, so there is no cost to keeping the option alive. However, if the dividend rate is positive, there is an opportunity cost to keeping the option alive rather than exercising it. That opportunity cost is the dividend stream that one forgoes by holding the option rather than the stock. Since δ is a proportional dividend rate, the higher is the price of the stock, the greater is the flow of dividends. At some high enough price, the opportunity cost of foregone dividends becomes great enough to make it worthwhile to exercise the option.

For our investment problem, μ is the expected rate of return from owning the completed project. It is the equilibrium rate established by the capital market, and includes an appropriate risk premium. If $\delta > 0$, the expected rate of capital gain on the project is less than μ. *Hence δ is an opportunity cost of delaying construction of the project, and instead keeping the option to invest alive.* If δ were zero, there would be no opportunity cost to keeping the option alive, and one would never invest, no matter how high the NPV of the project. That is why we assume $\delta > 0$. On the other hand, if δ is very large, the value of the option will be very small, because the opportunity cost of waiting is large. As $\delta \to \infty$, the value of the option goes to zero; in effect, the only choices are to invest now or never, and the standard NPV rule again applies.

The parameter δ can be interpreted in other ways. For example, it could reflect the process of entry and capacity expansion by competitors. (However, in Chapter 8 we will discuss more complete models that endogenize the process of rivals' entry, and find that the resulting equilibrium *cannot* be well described by simply raising the δ for each firm.) Or it can simply reflect the cash flows from the project. If the project is infinitely lived, then equation (1) can represent the evolution of V during the operation of the project, and δV is the rate of cash flow that the project yields. Since we are assuming that δ is rate r replaces the discount rate ρ. The same boundary conditions (10)–(12) will also apply here, and for the same reasons as before. Thus the solution for $F(V)$ again has the form

$$F(V) = AV^{\beta_1},$$

except that now r replaces ρ in the quadratic equation for the exponent β_1, and therefore

$$\beta_1 = \tfrac{1}{2} - (r-\delta)/\sigma^2 + \sqrt{\left[(r-\delta)/\sigma^2 - \tfrac{1}{2}\right]^2 + 2r/\sigma^2}. \tag{24}$$

The critical value V^* and the constant A are again given by equations (14) and (15).

Hence the contingent claims solution to our investment problem is equivalent to a dynamic programming solution, under the assumption of risk neutrality (that is, the discount rate ρ is equal to the risk-free rate).[9] Thus whether or not spanning holds, we can obtain a solution to the investment problem, but without spanning, the solution will be subject to an assumed discount rate. In either case, the solution will have the same form, and the effects of changes in σ or δ will likewise be the same. One point is worth noting, however. Without spanning, there is no theory for determining the "correct" value for the discount rate ρ (unless we make restrictive assumptions about investors' or managers' utility functions). The CAPM, for example, would not hold, and so it could not be used to calculate a risk-adjusted discount rate in the usual way.

4 Characteristics of the Optimal Investment Rule

Let us assume that spanning holds, and examine the characteristics of the optimal investment rule and the value of the investment opportunity, as given

[9] This result was first demonstrated by Cox and Ross (1976). Also, note that equation (23) is the Bellman equation for the maximization of the net payoff to the risk-free portfolio that we constructed. Since the portfolio is risk-free, the Bellman equation for that problem is

$$r\,\Phi\,dt = -\delta V F'(V)\,dt + \mathcal{E}(d\Phi), \tag{i}$$

that is, the return on the portfolio equals the per-period cash flow that it pays out [which is negative, since $\delta V F'(V)$ must be paid in to maintain the short position], plus the expected rate of capital gain. By substituting $\Phi = F - F'(V)V$ and expanding dF as before, one can see that equation (23) follows from (i). Also, note that in equation (i), $\delta = \mu - \alpha$ and not $r - \alpha$, so one must still have an estimate of the risk-adjusted expected return that applies to V. This is an example of the "equivalent risk-neutral valuation" procedure discussed in Chapter 4, Section 3.A.

by equations (13), (14), (15), and (24). Some numerical solutions will help to illustrate the results and show how they depend on the values of the various parameters. As we will see, these results are qualitatively the same as those that come out of standard option pricing models.

Unless otherwise noted, in what follows we set the cost of the investment, I, equal to 1, $r = 0.04$, $\delta = 0.04$, and $\sigma = 0.2$ (at annual rates). (Note that we do not need to know μ or α, but only the difference between them, δ.) Payout rates on projects vary enormously from one project to another, so this value of 4 percent for δ should be viewed as reasonable, but not necessarily representative. As for σ, the standard deviation of the rate of return on the stock market as a whole has been about 20 percent on average. Although this represents a diversified portfolio of assets, it also includes the effects of leverage on equity returns, and so might be a reasonable number for an average asset.

Given these parameter values, $\beta_1 = 2$, $V^* = 2I = 2$, and $A = \frac{1}{4}$. Thus the simple NPV rule, which says that the firm should invest as long as V is at least as large as I, is grossly in error. For this reasonable set of parameter values, V must be at least *twice* as large as I before the firm should invest. The value of the firm's investment opportunity is $F(V) = \frac{1}{4}V^2$ for $V \le 2$, and $F(V) = V - 1$ for $V > 2$ (since the firm exercises its option to invest and receives the net payoff $V - 1$ when $V > 2$).

Figure 5.3 plots $F(V)$ as a function of V for these parameter values, and also for $\sigma = 0$ and $\sigma = 0.3$. In each case, the tangency point of $F(V)$ with the line $V - I$ gives the critical value V^*. The figure also shows that the simple NPV rule must be modified to include the opportunity cost of investing now rather than waiting. That opportunity cost is exactly $F(V)$. When $V < V^*$, $F(V) > V - I$ and therefore $V < I + F(V)$: the value of the project is less than its *full* cost, the direct cost I plus the opportunity cost $F(V)$. [When $\sigma = 0$, $V^* = I$, and $F(V) = 0$ for $V \le I$.]

Note that $F(V)$ increases when σ increases, as does the critical value V^*. Thus greater uncertainty increases the value of a firm's investment opportunities, but (for that very reason) decreases the amount of actual investing that the firm will do. As a result, when a firm's market or economic environment becomes more uncertain, the market value of the firm can go up, even though the firm does less investing and perhaps produces less.

The dependence of V^* on σ is also shown more directly in Figure 5.4. Observe that V^* increases sharply with σ. *Thus investment is highly sensitive to volatility in project values, irrespective of investors' or managers' risk preferences, and irrespective of the extent to which the riskiness of V is correlated with the market.* Firms can be risk neutral, and stochastic changes in V can

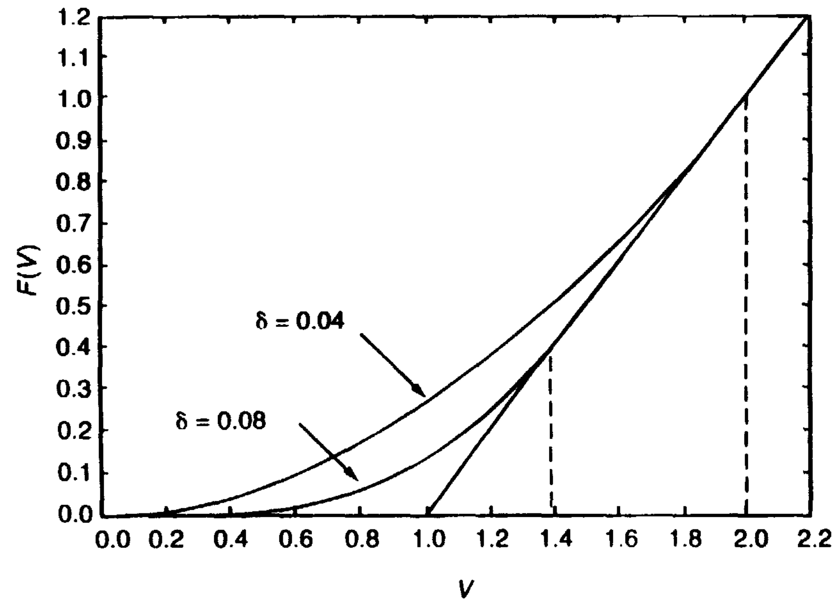

Figure 5.5. Value of Investment Opportunity, $F(V)$, for $\delta = 0.04$ and 0.08

than in the standard model. In the standard model, an increase in the interest rate reduces investment by raising the cost of capital; in this model, it increases the value of the option to invest and hence increases the opportunity cost of investing now. (Figure 5.7 shows the dependence of V^* on r for δ equal to 0.04 and 0.08.)

Once again in this calculation we held δ fixed as r increased. If instead we hold α fixed, then δ increases one for one with r. Now a lower r reduces β_1 and increases the critical level V^*. In this sense, a lower interest rate *discourages* investment. This is a pure manifestation of the option idea: a low interest rate makes the future relatively more important, therefore it increases the opportunity cost of exercising the option to invest.

Figure 5.8 provides another way of seeing how the optimal investment rule depends on the parameter values. It also lets us cast our results in terms of Tobin's q. Here we use the "value of assets in place" definition that ignores the opportunity cost of exercising the option, as explained in Section 2(c) above. Then $q^* = V^*/I = \beta_1/(\beta_1 - 1)$ is the critical value of this q, that is, the multiple of I required to invest. The figure shows contours of constant q^* plotted for different values of the parameter combinations $2r/\sigma^2$ and $2\delta/\sigma^2$.

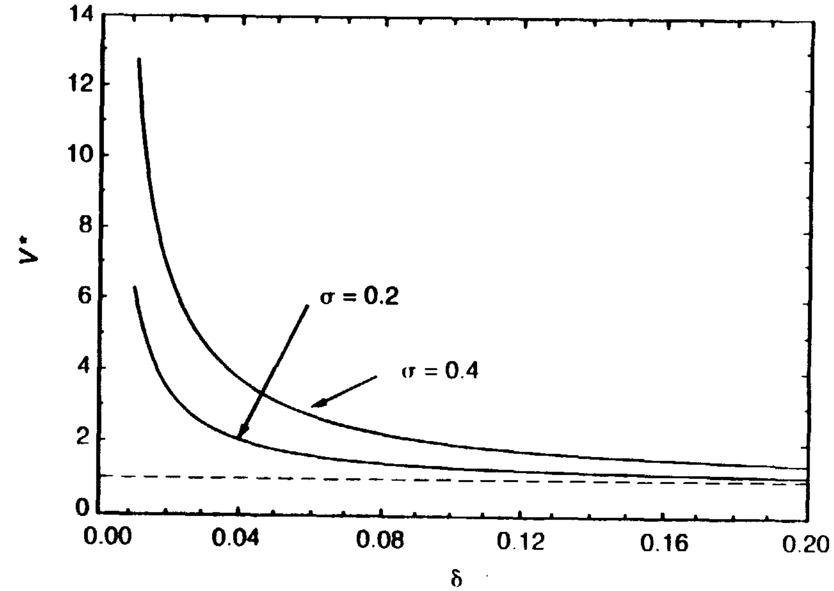

Figure 5.6. Critical Value V^* as a Function of δ

We have scaled r and δ by $2/\sigma^2$ because, as the reader can verify by substituting $\beta_1 = q^*/(q^* - 1)$ into equation (16), q^* must satisfy

$$\frac{2r}{\sigma^2} = q^* \left(\frac{2\delta}{\sigma^2}\right) - \frac{q^*}{q^* - 1}.$$

As the figure shows, the multiple is large when δ is small or r is large.

These comparative statics results are the same as those that apply to financial call options. Our option to invest is analogous to a perpetual call option on a dividend-paying stock, where V is the price of the stock, δ is the (proportional) dividend rate, and I is the exercise price of the option. The value of the call option on the stock and the optimal exercise rule will depend on the parameters σ, δ, and r as illustrated by Figures 5.1–5.7.[11]

We repeat that it is important to be careful when interpreting comparative statics results, because different parameters are unlikely to be independent of each other. For example, an increase in the risk-free rate, r, is likely to result

[11] For more detailed discussions of financial call options and their comparative statics, see Cox and Rubinstein (1985) and Hull (1989).

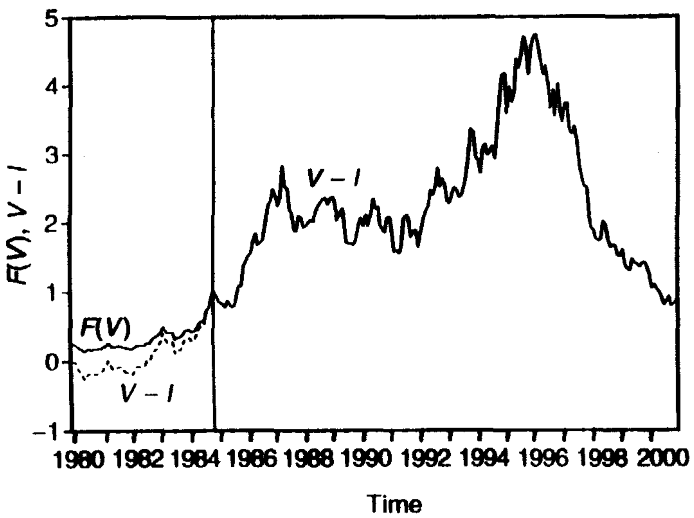

Figure 5.9. Sample Path of F(V) and V − I

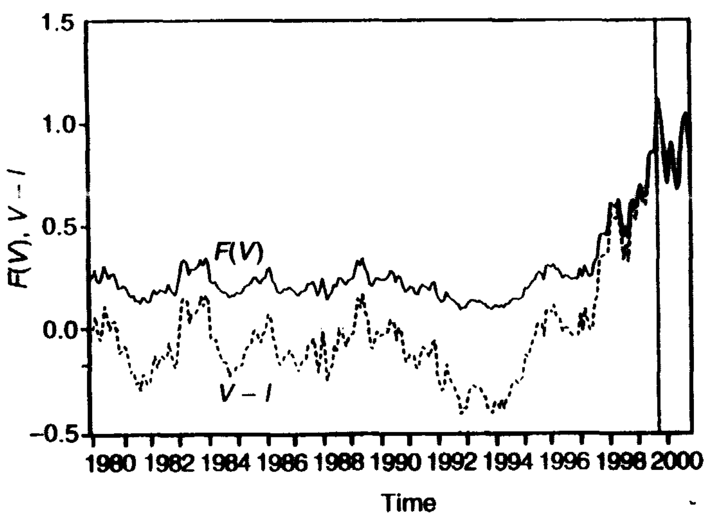

Figure 5.10. Another Sample Path of F(V) and V − I

where at each time t, ϵ_t is drawn from a normal distribution with zero mean and unit standard deviation. (Note that the coefficient $0.0577 = 0.20/\sqrt{12}$ is the *monthly* standard deviation.)

Since $V_0 = I = 1$, the standard NPV rule would call for investing immediately. However, $F(V_0) = 0.25$, so $V_0 < I + F(V_0)$, and the firm should wait rather than invest. In Figure 5.9, the firm happens to wait approximately five years before V reaches $V^* = 2$. This waiting time can vary considerably from one sample path to the next. In the sample path shown in Figure 5.10 for example, the firm must wait much longer—nearly 20 years—before V reaches the critical value of 2.[13]

5 Alternative Stochastic Processes

The use of a geometric Brownian motion as a model for V is convenient, but in some cases may not be realistic. In this section we will examine the value of the investment opportunity and the optimal investment rule when V follows alternative stochastic processes. We will first consider the case of a mean-reverting process, and then a Poisson jump process.

5.A Mean-Reverting Process

Suppose V follows the mean-reverting process

$$dV = \eta(\bar{V} - V)V\,dt + \sigma V\,dz, \tag{26}$$

so that the expected percentage rate of change in V is $(1/dt)\mathcal{E}(dV/V) = \eta(\bar{V} - V)$, and the expected absolute rate of change is $(1/dt)\mathcal{E}(dV) = \eta\bar{V}V - \eta V^2$, a parabola that equals zero at $V = 0$ and $V = \bar{V}$ and has a maximum at $V = \bar{V}/2$. As we will see, an advantage of this particular process is that we will be able to obtain an analytical solution to the investment problem.

To find the optimal investment rule, we will use contingent claims analysis. Let μ be the risk-adjusted discount rate for the project (that is, μ reflects the systematic risk in the stochastic fluctuations in V). In this case the expected

[13] The expectation and variance of this "waiting time" can be computed analytically. We will not need these expressions, but refer the interested reader to some simple cases in Dixit (1993a, pp. 54–57), and the more rigorous theory in Karlin and Taylor (1981, pp. 242–244) or Harrison (1985, pp. 11–14).

Ch. 2

Investment under Uncertainty

Avinash K. Dixit and Robert S. Pindyck

Princeton University Press
Princeton, New Jersey

In practice, most firms do not enjoy monopoly rights to invest, but instead must consider the possible entry of new competitors, or expansion of existing ones. This raises a fundamental doubt concerning our earlier conclusions. The opportunity to wait, and its value, depend on what the firm's competitors do. With free entry, should this value not be reduced to zero? Would that not restore the Marshallian criteria comparing price to the long-run average cost in the case of investment, and to the average variable cost for disinvestment? Thus the reader might suspect that the theory of an individual firm would not survive an extension of the scope of the analysis to the level of industry equilibrium.

In this chapter and the following one, we take up these questions. The answers are largely reassuring. What happens to the value of waiting depends not only on the nature of competition, but also on the nature of the uncertainty. We find that when uncertainty is firm-specific, a firm's value of waiting survives, and our firm-level analysis can readily be extended to the industry equilibrium. For aggregate or industry-wide uncertainty, the value of waiting for any one firm does drop to zero, but that does not restore the Marshallian criteria of price-cost comparisons. The optimal investment and disinvestment threshold prices differ from costs just the same way as they did in our firm-level analysis of Chapters 5–7, albeit for a different reason. We can no longer specify the stochastic process of the price exogenously. Price is an endogenous variable of industry equilibrium. We must trace the uncertainty to a deeper level, namely, the demand and cost conditions. The endogenous feedback of new entry on price is what generates the gap between the Marshallian and the optimal rules when uncertainty is industry-wide.

The earlier analysis of the monopoly firm also serves a useful pedagogical purpose, as well as a substantive one. All of our industry-level analysis is conducted using the same techniques (dynamic programming and contingent claims analysis) that were developed and illustrated in the preceding chapters. Thus that work provides a simpler setting for readers to become familiar with the new methods.

This chapter concerns the basic theory of industry equilibrium. In other words, we take the firm's decision model of Chapters 5–7, and build it as directly as possible into a model of industry equilibrium. Then in the next chapter we take up various extensions and implications. We allow for heterogeneity among firms, and we consider a simple example of imperfect competition among a small number of firms in the industry.

Finally, and perhaps most importantly, in Chapter 9 we also examine issues of policy regarding investment. If firms' irreversible choices under uncertainty are significantly influenced by the option value of the status quo and therefore characterized by considerable inertia, should the government attempt to encourage investment? How will various policy instruments affect investment? In particular, what will be the effect of government policies to reduce uncertainty (for example, through the use of price controls)? What will be the effect of uncertainty concerning the government's own actions (for example, uncertainty over future tax rates and regulatory changes)? Such questions must be examined at the industry level if they are to be useful guides to policy in many practical situations; therefore these chapters are the right place for their study.

Before we start on this program, we should say a few words about the nature of irreversibilities in relation to industry-wide and firm-specific forms of uncertainty. Investment is partially or totally irreversible when some or all of its costs are sunk. In Chapter 1, Section 3 we offered some general reasons why this happens. Perhaps the most prominent was the specificity of the plant and equipment itself. This applies with greater force in the case of industry-wide uncertainty than for firm-specific uncertainty. A steel plant cannot be used outside the steel industry. If one steel firm suffers an idiosyncratic negative shock, it can sell its plant to another firm and get fairly good value for it, so the irreversibility is less severe. However, if the whole industry suffers a negative shock, then the resale value of the plant is small and the irreversibility is large. Thus we should expect that our theory has greater significance in the context of aggregate uncertainty. Of course, even for firm-specific shocks, some investment expenditures are sunk, for example, any research or exploration costs incurred in the process of discovering the firm-specific random shock. Also, even nonspecific capital (such as automobiles, computers, and office equipment) is subject to a loss in resale value due to asymmetric information about product quality [that is, the "lemons" problem, illustrated by Akerlof (1970)].

1 The Basic Intuition

Before turning to the mathematical models, we enlarge on the intuition for the results to come. First suppose the uncertainty is firm-specific. Thus different firms experience independent shocks to demand (for example, a shift of fashion in an industry with differentiated products) or cost (for example, a chance improvement in entrepreneurial skills). As in the previous chapters, suppose each firm's shock has positive serial correlation; in actual models we will specify these shocks as Brownian motions. Even though the firms are identical ex ante, a firm that experiences a favorable shock does sneak a lead other firms enter, the industry supply curve shifts to the right, and the price rises less than proportionately with Y. Therefore price is a concave function of Y, and then so is the profit flow. Greater uncertainty in Y now reduces the expected value of investing relative to that of not investing. That is why the firm requires a higher current profitability (in excess of the Marshallian normal return) before it will invest.

We should stress the similarity as well as the difference between the two scenarios. In each, the underlying symmetric demand shock translates into an asymmetric profit flow shock; but this happens in very different ways in the two cases. In the case of firm-specific uncertainty, the downside of the profit shock is cushioned by the possibility of waiting. Thus greater uncertainty makes waiting more valuable relative to investing at once. In the case of industry-wide uncertainty, any one firm in the mass of competitive potential firms has a zero value of waiting. However, the upside profit potential is cut off by the prospect of entry of other firms. Therefore greater uncertainty reduces the value of investing relative to that of not investing at all.

In reality there are several other factors that can affect the convexity or concavity of profit flows as a function of the underlying shock variable. If the firm can adjust some variable inputs instantaneously, then its profit flow becomes a convex function of the price, as we saw in Chapter 6, Section 3. In addition, in Chapter 11 we will see that when the firm can add to its capital stock, and its output flow is given by a production function, there can be other ways by which the marginal profitability of an incremental investment becomes convex in price. However, the above intuition still operates similarly. For example, with firm-specific uncertainty, the possibility of waiting cuts off the downside risk and makes the profit flow an *even more* convex function of the underlying shock variable. In most of Chapters 8 and 9, however, we will leave aside these additional sources of convexity. We will define each firm as the possessor of a technology to install and operate a single discrete project of fixed size, and focus on the two kinds of asymmetries explained just above.

2 Aggregate Uncertainty

Despite the simplicity of the underlying intuition, the general model of uncertainty is quite difficult to set up and solve formally. The ideas are easier to explain by constructing some special cases that together span the complexity of the general one, so that is how we will proceed. In this section and the next, we consider only the industry-wide shock Y. Then, in Section 4, we deal with purely firm-specific uncertainty. Finally, some insights that do depend

in an essential way on the joint presence of the two kinds of uncertainty are examined in the context of a simple but general model in Section 5.

When all uncertainty is industry-wide, the multiplicative factor X in the general demand curve (1) is constant, so we can just set it equal to 1. Then the industry's inverse demand curve becomes

$$P = Y\,D(Q). \tag{2}$$

The aggregate shock Y will follow the geometric Brownian motion process

$$dY = \alpha\, Y\, dt + \sigma\, Y\, dz. \tag{3}$$

On the production side, we assume that there is a large number of risk-neutral competitive firms. Each firm can undertake a single irreversible investment, requiring an initial sunk cost I. Once this investment is made, it yields a flow of one unit of output forever with no variable cost of production. We embed such firms in an industry by supposing that each unit of output is very small relative to the total industry output Q, so that each firm is an infinitesimal price taker. When Q firms are active, the short-run equilibrium price can be determined from equation (2) above.

As we discussed before, this is the simplest continuation of the model of Chapters 5–7 that serves our present purpose. Later and in Chapter 9, we introduce various generalizations, where each firm has some variable cost, short-run output variability, exit possibilities, etc., where the shocks affect demand in more general ways, and where the industry has some imperfect competition.

To set the stage for the competitive industry equilibrium, think of the usual textbook static model. The industry price—a single number—is parametric to each firm. The sum of the individual firms' optimum quantity responses to the price constitutes the industry's supply function. The equilibrium price is determined by the condition equating industry demand and supply. In our dynamic world with uncertainty, the corresponding equilibrium concept is one of rational expectations. Each firm takes as exogenous the whole stochastic process of the price. So we start with a price process, let all firms respond to it, and then find the process that clears the market at each instant. This is a function or a mapping that takes us from one stochastic process to another. We have an equilibrium if we get the same price process that we started with, or in other words, a *fixed point* of the mapping.

Since a stochastic process as a whole is a complex mathematical object—a vector in an uncountably infinite-dimensional function space—finding such a fixed point in full generality is far too difficult. Luckily, the solution to our problem can be found using a much simpler method.

This looks like a smooth-pasting condition, but it is not a consequence of any optimization. Such a condition holds at any reflecting barrier for a diffusion process.[6]

Now we can solve for B from equation (7) to get

$$B = -\overline{P}^{1-\beta_1} / (\beta_1 \delta).$$

Note that $B < 0$; as explained earlier, the barrier cuts off some upside price potential, so the correction to the value is a reduction. Substituting for B into (6), we have

$$v(P) = \frac{P}{\delta} - \frac{1}{\delta\beta_1} P^{\beta_1} \overline{P}^{1-\beta_1}. \tag{8}$$

2.B Equilibrium

The quick way to find the industry's equilibrium is to use a dynamic zero excess profit condition. At $\overline{P}$, the common entry threshold for all firms, each firm is just indifferent between entering and staying out, so the value of being in, $v(\overline{P})$, must exactly equal the entry cost I. Using equation (8) above, this gives

$$\overline{P} = \frac{\beta_1}{\beta_1 - 1} \delta I. \tag{9}$$

Most remarkably, this is the same entry price as that for a unit-sized monopolist firm facing the same demand process; compare this to equation (9) of Chapter 6. The two situations differ in two ways. The monopolist of Chapter 6 was not threatened by entry, so there was no upper barrier on the price process; now there is. However, the monopolist had a positive option value of waiting, while any of several identical potential firms of this chapter must have zero value of waiting. It so happens that the two differences exactly offset each other.

This coincidence between a competitive firm's and a monopolist's entry threshold is in the context of a very special example. In Chapter 9 we will find a very general result of this kind—a competitive firm can make the correct investment decision by acting myopically in the matter of future competitive entry, and acting as if it were going to be the last firm ever to enter this industry. That result also rests on a similar exact offset of two effects, one on the value of investing and the other on the value of waiting.

[6]See Malliaris and Brock (1982, p. 200) or Dixit (1993a, Section 3.5).

To understand the solution more fully, we must go into the details of the fixed-point process for constructing the equilibrium. Consider a firm contemplating entry. Write $f(P)$ for the value of its option to enter. As in Chapter 6, this takes the form

$$f(P) = A P^{\beta_1},$$

where A is a constant to be determined, and β_1 is as above. If the firm decides to enter when the price is P, it pays the investment cost I and receives in return an asset that we just valued at $v(P)$. The optimal entry threshold P^* satisfies two familiar conditions. First, value matching:

$$f(P^*) = v(P^*) - I,$$

and second, smooth pasting:

$$f'(P^*) = v'(P^*).$$

Using the functional forms for the functions $f(P)$ and $v(P)$, we have

$$A(P^*)^{\beta_1} = B(P^*)^{\beta_1} + P^*/\delta - I,$$

and

$$\beta_1 A(P^*)^{\beta_1 - 1} = \beta_1 B(P^*)^{\beta_1 - 1} + 1/\delta.$$

Note that we have already solved for the constant B in terms of the assumed upper barrier $\overline{P}$, but some expressions convey more insight when B is retained as such.

These two equations can be solved for the threshold P^* and the constant A; we have

$$P^* = \frac{\beta_1}{\beta_1 - 1} \delta I, \tag{10}$$

and

$$\begin{aligned} A &= B + \frac{1}{\beta_1 \delta} (P^*)^{1-\beta_1} \\ &= \frac{1}{\beta_1 \delta} \left[(P^*)^{1-\beta_1} - \overline{P}^{1-\beta_1} \right]. \end{aligned} \tag{11}$$

Observe two features of the solution: the barrier $\overline{P}$ affects the solution only via the constant B in the value function $v(P)$, and the constant A in the option value function $f(P)$ responds one for one to changes in B. These have important implications for the equilibrium.

it will earn a normal return on its sunk cost. However, we saw in Chapter 6 that it does not invest until the price rises to P^*, which is $\beta_1/(\beta_1 - 1)$ times P_0. We explained this in terms of the option value of waiting.

Now we see that a unit-sized competitive firm also waits until the price rises to the same level, even though its option value of waiting is zero. The explanation lies in the difference between the price processes in the two cases. The competitive firm's price process has an upper barrier, which reduces its expectation of future prices and returns. Specifically, since the firm knows that all other firms face the same choice and make the same decisions, the price will never rise above the level that prevails at its instant of entry; the current price when it enters is not the average but the best price it will ever get. If competitive firms adopted the rule of entering when the price reached P_0, they would earn a normal return only at those instants when entry was taking place—they would earn lower returns at all other instants. The average return over time would then be insufficient to justify the initial investment expenditure. On the other hand, when the entry threshold exceeds P_0, each firm will experience some period of supernormal returns and some periods of subnormal returns. The equilibrium $\overline{P}$ is exactly the level that ensures a normal return on average.

Since the entry threshold cum industry equilibrium price coincides with the threshold for a monopoly with the same parameters α, σ, r, and δ, we need not present detailed numerical calculations for the competitive equilibrium case; instead we refer the reader to those in Chapters 5 and 6. However, a few summary numbers are useful. Table 8.1 shows β_1 and the current rate of return on investment at the threshold,

$$\overline{P}/I = \beta_1\delta/(\beta_1 - 1),$$

for $r = 0.05$, $\alpha = 0$ and 0.03, and $\sigma = 0, 0.2$, and 0.4. Note that when $\alpha = \sigma = 0$, the return $\overline{P}/I$ equals the Marshallian return, that is, the interest rate $r = 0.05$. This is also the case when $\alpha = 0.03$ and $\sigma = 0$. [As discussed in Chapter 5, when α is positive there is a value to waiting even if there is no uncertainty, and indeed in this case $\beta_1/(\beta_1 - 1) = 2.5$, but since δ falls as α increases, the return remains equal to r.] For either value of α, as σ is increased to 0.2 and 0.4, β falls and the required return $\overline{P}/I$ rises to about two or three times its Marshallian value. Hence the general finding from Chapters 5 and 6, that the firm's optimal decisions differ substantially from the implications of the textbook present value approach, has an exact parallel for a competitive industry. Its equilibrium differs substantially from the picture offered by the Marshallian theory presented in most elementary and intermediate microeconomics textbooks.

Table 8.1. Required Return for Competitive Entry
Note: $r = 0.05$

α	σ	β_1	$\overline{P}/I$
0	0	∞	0.050
0	0.2	2.16	0.093
0	0.4	1.44	0.165
0.03	0	1.67	0.050
0.03	0.2	1.35	0.077
0.03	0.4	1.16	0.143

3 Industry Equilibrium with Exit

The above basic model closely followed that of the monopoly firm in Chapter 6, and gave us very analogous results for the competitive industry with aggregate uncertainty. Most of the extensions of this model are left for Chapter 9. In this chapter we take up just one that fits more naturally here. We introduce exit, and construct a model that closely follows that of the monopoly firm's entry and exit decisions in Chapter 7. Once again, the results for the competitive industry with aggregate uncertainty are thoroughly parallel.

For exit to be a meaningful option, we need two conditions. First, the operating profit flow must sometimes become negative; we make this possible by introducing a variable cost C for each unit-sized firm. Second, temporary suspension of operation and resumption without a cost penalty must be ruled out; we do so. We also introduce a lump-sum cost of exit E. As before, this can comprise any legally required severance payments or costs of restoring land. It can also be negative (but numerically less than cost I), representing any nonsunk portion of the entry cost.

Now the intuition is that the exit of other firms will generate a floor—a lower reflecting barrier—on the price process, just as their entry generated a ceiling—an upper reflecting barrier. Each firm will have rational expectations about the price process it faces, namely, a geometric Brownian motion between these two barriers. The firm's own entry and exit decisions will again take the form of upper and lower thresholds on the price. The equilibrium levels of the two barriers will be found from a fixed-point argument: each firm's thresholds should equal the barriers generated by the behavior of all firms in the industry.

Regard these as a pair of linear equations in A_1, A_2. The coefficient matrix

$$\begin{pmatrix} \beta_1 \overline{P}^{\beta_1 - 1} & \beta_2 \overline{P}^{\beta_2 - 1} \\ \beta_1 \underline{P}^{\beta_1 - 1} & \beta_2 \underline{P}^{\beta_2 - 1} \end{pmatrix}$$

is nonsingular as long as $\overline{P} > \underline{P}$. Then the only solution is $A_1 = A_2 = 0$. Therefore the value of an idle firm is identically zero, as it should be, given competitive conditions and identical firms. This completes the solution.

3.A Entry, Exit, and Price in the Copper Industry

We now return to our example of entry and exit in the copper mining industry from Chapter 7, restating some of the numerical results to give the readers a better feel for the magnitudes involved. In fact, the use of industry data to illustrate the story of a firm having a monopoly right to invest was something of an anomaly in Chapter 7; now the same numbers have a more satisfactory interpretation in the context of industry equilibrium. In the central case studied, we assumed that the capital cost of building an average-sized mine, smelter, and refinery (producing 10 million pounds of copper per year) was $I = \$20$ million, and the cost of site restoration upon abandonment was $E = \$2$ million. The variable cost was $C = \$0.80$ per pound, but was allowed to vary around this figure. The price volatility parameter σ was 0.2 in annual units, and was also allowed to vary around this range. The riskless interest rate was $r = 0.04$, and the return shortfall was $\delta = 0.04$. With these numbers, the Marshallian entry threshold price would be \$0.88 and the exit threshold \$0.792. As shown in Table 8.2, however, the correct entry and exit thresholds are \$1.35 and \$0.55, respectively. The table also shows the Marshallian and the actual thresholds for other values of C and σ.

Figures 8.2 and 8.3 show, for the central case of $C = \$0.80$ per pound and $\sigma = 0.2$, sample paths for the price of copper. Observe that this price fluctuates as a geometric Brownian motion between the upper and lower reflecting barriers, which are the entry and exit thresholds of \$1.35 and \$0.55. These thresholds are shown in the figures as horizontal lines; also shown are the Marshallian thresholds of \$0.88 and \$0.79. (In the figures, time is measured in years, and each year was divided into 50 increments for purposes of generating sample paths.) Figure 8.2 shows a "fortunate" (from the point of view of a copper producer) sample path, in which the price spends much of the time at the upper end of its range, while Figure 8.3 shows an "unfortunate" realization.

These figures show particular realizations of price, but we could ask what percentage of the time the price should be expected to stay in different regions

Table 8.2. Entry and Exit Thresholds in Copper Mining (Note: see text for parameters)

		Upper threshold		Lower threshold	
c	σ	Marshallian	Correct	Marshallian	Correct
0.8	0.1	0.88	1.12	0.792	0.63
0.8	0.2	0.88	1.35	0.792	0.55
0.8	0.4	0.88	1.75	0.792	0.45
0.4	0.2	0.48	0.80	0.392	0.26
0.6	0.2	0.68	1.06	0.592	0.40
0.8	0.2	0.88	1.35	0.792	0.55
1.0	0.2	1.08	1.60	0.992	0.70

of its range over the long run. We can answer this question by calculating the long-run stationary distribution for price. Begin by observing that P follows the geometric Brownian motion of equation (4) between its reflecting barriers $\underline{P}$ and $\overline{P}$. Therefore, using Ito's Lemma, we know that $p \equiv \log P$ follows a simple Brownian motion with the drift parameter $\alpha' = \alpha - \frac{1}{2}\sigma^2$ and the variance parameter σ, between corresponding reflecting barriers $\underline{p} = \log \underline{P}$ and $\overline{p} = \log \overline{P}$. Now we can use the result of Chapter 3, Section 5 to find the long-run distribution of p. It is an exponential distribution, with density $K e^{\gamma x}$, where $\gamma = 2\alpha'/\sigma^2$, and the constant of proportionality K is chosen to achieve a total probability mass of unity. With this, it is easy to calculate what proportion of time p spends in various subsets of the range $(\underline{p}, \overline{p})$. The results can then be translated into the corresponding ranges for P.

For our base case parameter values, we find that, on average, price will be between the upper Marshallian and entry thresholds of \$0.88 and \$1.35 about 58.5 percent of the time. Thus more than half the time, copper producers will be earning what in traditional microeconomic analysis we would call a supernormal profit. Price will be between the two Marshallian thresholds of \$0.79 and \$0.88 about 11.3 percent of the time (in which case we can say that firms are earning positive but subnormal profits). Finally, price will be between the exit threshold of \$0.55 and the lower Marshallian threshold of \$0.79, so that firms are incurring losses, about 30.2 percent of the time.

These figures show very dramatically how the dynamics of a competitive industry under conditions of uncertainty will differ from the textbook picture. For almost 90 percent of the time, we expect the copper industry to be in a

An example will help fix the idea. Consider a pharmaceutical company that can develop a new drug by incurring the research cost R. This yields an initial estimate of its efficacy and profitability. The firm patents the drug, but unless the profit estimate is sufficiently high, it will not incur the additional investment expenditure I that is necessary to begin production. Over time the profit estimate may increase as new uses are found for the drug, or decrease as other drugs to treat the same condition are discovered by other firms.

We characterize the competitive equilibrium of such an industry in the long run. There are numerous competitive firms facing independent shocks, and there is substantial uncertainty and volatility at the firm level. However, different firms' shocks are independent, and the operation of the law of large numbers ensures that industry aggregates are nonrandom.[8] Thus a nonrandom total volume of output can be produced by firms whose identities change through time but whose aggregate population distribution remains stationary. However, the firm-level uncertainty leaves a mark on the industry equilibrium: the parameters of the distribution of active firms, and therefore the actual values at which the nonrandom industry quantity and price settle, do depend on the extent of uncertainty faced by each firm.

The idea that relatively tranquil industry-wide conditions conceal much firm-level uncertainty has been emphasized in recent empirical work. Davis and Haltiwanger (1990) and others have demonstrated quite impressively the large gross hirings and firings that underlie small net changes in employment in the U.S. economy. The models that are constructed for applications of this kind generally contain too much context-specific detail to let the general intuition stand out. Our simple model can help the reader develop a better conceptual understanding and more general intuition for such phenomena.

We begin by specifying the nature of uncertainty in X so as to fit with the firm's two stages of decisions. A new entrant gets an initial draw of X from a known distribution. Thereafter its X evolves as a geometric Brownian motion

$$dX = \alpha\, X dt + \sigma\, X dz. \tag{26}$$

We have interpreted X as an idiosyncratic demand shock (random fluctuations of taste shifts giving rise to price premia for slightly different varieties in the industry). What ultimately matters is the shock to profitability, and we could also think of X as a technology shock that appears in a reduced form in the formula for the firm's profit flow after the instantaneously variable choices have been optimized out, as discussed in Chapter 6, Section 3.

[8] Recall that we are not giving a formal rigorous treatment of this.

There is free entry into the industry, and anyone can get the initial draw by paying R. However, there is no obligation to start production at once. A further sunk investment I must be incurred to activate the process, and the firm can wait to see if X evolves to a more favorable level before making this irreversible commitment.

We will characterize the long-run stochastic equilibrium of such an industry. In fact we will postulate such an equilibrium, and then determine endogenously the various stationary magnitudes (price, number of firms, etc.) that constitute the equilibrium. Suppose N is the nonrandom stream of new entrants who pay the fee R and learn their initial X. Then their shocks will evolve independently and stochastically. A nonrandom flow M will reach the activation decision. We also want to keep the total number of active firms, Q, constant. To permit this, we assume that all firms, whether waiting or active, face an exogenous Poisson process of death with parameter λ. This process is also independent across firms. Then in a stationary equilibrium M must equal $\lambda\, Q$.

All uncertainty being idiosyncratic, we specify that each firm is risk neutral and makes its decisions to maximize its expected net worth. Let r denote the risk-free interest rate at which future profit flows are discounted.

4.A The Activation Decision

In the long-run stationary equilibrium with a large and constant number Q of active firms, each new entrant or waiting firm takes this Q as given. Its profit flow is $X\, D(Q)$. It continually observes X, and decides when to pay its investment cost I and become an active producer. This is formally identical to the basic single-firm model we studied in Chapter 6, Section 1. Equation (9) of that section gave us the price threshold P^* that triggered investment. In the notation of this chapter, that becomes a threshold X^* on the firm's shock, and the defining equation becomes

$$X^*\, D(Q) = \frac{\beta_1}{\beta_1 - 1}\,(r + \lambda - \alpha)\, I, \tag{27}$$

where β_1 is the positive root of the fundamental quadratic

$$\mathcal{Q} \equiv \tfrac{1}{2}\sigma^2 \beta(\beta - 1) + \alpha\,\beta - (r + \lambda) = 0.$$

The condition that ensures convergence of the expected profit flow is $r + \lambda > \alpha$; note that the Poisson death probability acts like a discount rate in achieving convergence, as we saw in the discussion of Poisson processes in Chapter 3.

Even now the argument is not complete. The number of active firms Q arises from a complex chain of initial entry decisions, independent random fluctuations of the firms' shock variables X, subsequent entry decisions, and independent random deaths. We must show how these interact in a consistent way to produce the industry's equilibrium Q.

4.C The Distribution of Firms

Recall the actual life history of any one firm that has just paid the entry cost R. It begins with an X randomly drawn from its known distribution. If the initial X exceeds the threshold X^*, the firm pays the investment cost I and becomes an active producer at once. Otherwise it lets its X evolve, and activates if and when X^* is reached. Throughout this process, the firm faces a constant and exogenous probability rate λ of death.

Such new entrants arrive at rate N. The full stochastic dynamics of each of them—the probability that it will be alive and occupy a position X at time t—can be examined using the Kolmogorov equation, which we developed in the Appendix to Chapter 3. Here our aim is more limited. For industry equilibrium, only the total numbers of firms in various states matter—how many are active, and how many are waiting with what values of X. Therefore the law of large numbers allows us to restrict attention to a long-run stationary equilibrium. This means that the rates of Poisson death by exit, and of activation, are constant through time. Likewise, the numbers of firms with various current levels of X are constant through time. Of course the actual identities of the firms occupying these positions keep changing, but for our purpose any firm is like any other with the same X.

The method of calculating this long-run distribution of firms is the same as that of Chapter 3, Section 5, but now we must include two new features, namely, fresh entry and Poisson deaths. It proves more convenient to work in terms of the logarithm, $x = \log X$. Let $g(x)$ denote the density function of the initial draw of x, and $G(x)$ the corresponding cumulative distribution. Note that the range of x extends to $-\infty$ to the left. Let $x^* = \log X^*$. Of the newly entering firms, $N[1 - G(x^*)]$ immediately get a draw large enough to justify activation. The rest join the mass of firms that do not complete the second step of committing the investment cost at once, but wait to reach the activation threshold.

For both groups, x continues to evolve. Applying Ito's Lemma to (26), we see that x follows the Brownian motion

$$dx = \nu\, dt + \sigma\, dz, \tag{31}$$

where $\nu = \alpha - \frac{1}{2}\sigma^2$. Also, both groups suffer exogenous "deaths" under the Poisson process with parameter λ.

Begin with the waiting firms, which are distributed over the range $(-\infty, x^*)$. Let $N\phi(x)$ denote the density of such firms at location x; the factor N just scales this by the rate of entry and leads to a simpler equation for $\phi(x)$. For the density to remain constant through time, the rate at which firms arrive at x (having received positive shocks from below or negative shocks from above) must equal the rate at which firms at x move away (having received shocks of the Brownian motion process or Poisson death). We express this equation of "balanced flow" of firms in a more precise way.

For this purpose, we use the binomial approximation to Brownian motion that proved so useful in Chapter 3, Section 2(b). Divide time into short intervals of duration dt, and the x space into short segments, each of length $dh = \sigma\sqrt{dt}$. Of the firms located in one such segment, in one short time interval a proportion $\lambda\, dt$ will die. Of the rest, a fraction p will move one segment to the right, and a fraction q will move to the left, where

$$p = \tfrac{1}{2}\left[1 + \frac{\nu}{\sigma}\sqrt{dt}\right], \qquad q = \tfrac{1}{2}\left[1 - \frac{\nu}{\sigma}\sqrt{dt}\right].$$

Now consider the segment centered at x. It starts out with $N\phi(x)\,dh$ firms. In the next unit time period dt, all of these move away with either Poisson or Brownian shocks. New entrants, as well as firms from the left and right, arrive to take their places. Figure 8.5 shows these flows schematically.

For balance we need

$$\begin{aligned} N\phi(x)\,dh = {} & N g(x)\,dh + p(1 - \lambda\,dt)\,N\phi(x - dh)\,dh \\ & + q(1 - \lambda\,dt)\,N\phi(x + dh)\,dh. \end{aligned} \tag{32}$$

Cancelling the common factor $N\,dh$, expanding the $\phi(x \pm dh)$ on the right hand side by Taylor's theorem, and simplifying, we get the differential equation

$$\tfrac{1}{2}\sigma^2\phi''(x) - \nu\phi'(x) - \lambda\phi(x) + g(x) = 0. \tag{33}$$

This equation is slightly different from the one of Chapters 5–7 because it pertains to a simple rather than geometric Brownian motion. However, the method of solution is very similar. It is easy to verify that the general solution has the form

$$\phi(x) = C_1 \exp[\gamma_1 x] + C_2 \exp[\gamma_2 x] + \phi_0(x).$$

The constant C is to be determined from the condition $\phi(x^*) = 0$. This yields

$$\phi(x) = \frac{1}{\lambda + \nu - \frac{1}{2}\sigma^2} \left[e^{x-\hat{x}} - e^{\gamma(x-x^*)} e^{x^*-\hat{x}} \right]. \tag{34}$$

From this we can calculate some aggregates for the waiting firms. Their total number is

$$M \equiv N \int_{-\infty}^{x^*} \phi(x)\,dx = \frac{N}{\lambda + \nu - \frac{1}{2}\sigma^2} \frac{\gamma}{\gamma - 1} e^{x^*-\hat{x}}.$$

The rate of activation is

$$-\tfrac{1}{2}\sigma^2 N \phi'(x^*) = N \frac{\frac{1}{2}\sigma^2(\gamma - 1)}{\lambda + \nu - \frac{1}{2}\sigma^2} e^{x^*-\hat{x}}.$$

We could similarly find the distribution of the mass of active firms. These will extend over the entire range $x \in (-\infty, \infty)$, because some firms may have activated and then had their X decline. They are also augmented by that part of the new entry flow that finds its initial X above X^* and activates immediately, and by the activation flow at x^*. Their numbers are diminished by the flow of Poisson deaths.

In fact we only need the total number of active firms, not the distribution of their X values, to find the industry equilibrium. The new entry flow that activates immediately is $N[1 - G(x^*)]$. The activation flow was found from the solution for waiting firms above as $-\frac{1}{2}\sigma^2\phi'(x^*)$. The population of active firms being Q, the death flow is λQ. To keep the numbers constant, we need

$$\lambda Q = N[1 - G(x^*) - \tfrac{1}{2}\sigma^2\phi'(x^*)]. \tag{35}$$

We can calculate x^* and Q from earlier conditions of equilibrium: the activation condition (27) and the free entry condition (30). Therefore this equation defines N. Since new entrants are indifferent to their choice because of the free entry condition (30), we can adopt the usual convention of long-run equilibrium analysis and suppose that just enough of them do enter. Thus the "stochastically stationary" equilibrium depicted above can be sustained.

In our special case where X has the uniform distribution, the condition (35) becomes

$$\lambda Q = N \left[1 - e^{x^*-\hat{x}} \frac{\lambda + \nu - \frac{1}{2}\sigma^2\gamma}{\lambda + \nu - \frac{1}{2}\sigma^2} \right].$$

This allows us to calculate the rate of new entry N, either relative to the mass of active firms Q, or relative to the total mass of active and waiting firms $(Q + M)$.

The various algebraic expressions get complicated, but some general principles stand out. When the firms in an industry are subject to specific shocks, the investment decisions of each are importantly influenced by the option value of waiting for better realizations of its own shock. At the industry level, the shocks and the responses of firms can aggregate into long-run stationary conditions, so that the industry output and price are nonrandom. However, the equilibrium levels of these variables are affected by the parameters of firm-specific uncertainty. Also, behind the aggregate certainty lies a great deal of randomness and fluctuations: firms enter, invest, and exit in response to the shocks to their individual fortunes.

In reality, industry-wide and firm-specific shocks occur together. Therefore we will proceed to combine the models studied thus far in this chapter into a single general model that encompasses both kinds of shocks. Then in Chapter 12 we will consider an application to some real data.

5 A General Model

Now we examine industry equilibrium in a more general model where the two kinds of uncertainty, firm-specific and industry-wide, coexist. To allow for the added notational and mathematical complexities of the joint uncertainty, we make some simplifications in each. Our treatment closely follows Caballero and Pindyck (1992).

We assume that the firms are risk neutral (or that the industry-wide risk is uncorrelated with the risk of the economy's capital market as a whole), leaving the more general case for the reader. We suppose that the industry demand is isoelastic; thus the demand equation (1) becomes

$$P = X\,Y\,Q^{-\epsilon}, \tag{36}$$

where $1/\epsilon$ is the price elasticity of demand. We also omit the second activation stage from the above model of firm-specific uncertainty; thus $I = 0$ and $X^* = 0$.

As before, each firm has the capacity to produce a unit of output and has no variable costs of production; thus, P is also its profit flow. Also as before, we can think of this as the reduced form of a more general structure in which any variable inputs (those without irreversibility or adjustment cost) are chosen at their optimal levels. We continue to treat firms as infinitesimal and their number as a continuous variable; then Q equals this number.

The actual value of a firm with a given X is $V(X, W) = X v(W)$. A prospective entrant, however, observes only W. Therefore the expected value from the entrant's perspective is $\overline{X} v(W)$. The threshold $\overline{W}$ is defined by the condition of free entry, so this expected value must equal the cost of entry R at that point. Substituting and simplifying, we find

$$\overline{X}\,\overline{W} = \frac{\beta_1}{\beta_1 - 1}\,(r + \lambda - \alpha_x - \alpha_w)\,R. \tag{43}$$

This is a very natural generalization that combines the corresponding equations for the case of pure industry-wide uncertainty—equation (10) of Section 2 above—and pure firm-specific uncertainty—equation (27) of Section 4 above. Its properties should by now be too familiar to need comment.

For further theoretical details of this model we refer the reader to Caballero and Pindyck (1992). In Chapter 12 we will discuss an empirical application of it.

6 Guide to the Literature

Lucas and Prescott (1971) considered the rational expectations equilibrium of investment in a competitive industry using a discrete-time Markov chain model. They established the optimality of the equilibrium. Lippman and Rumelt (1985) combined entry and exit in a similar model.

Edleson and Osband (1988) showed that a competitive firm's entry and exit thresholds in equilibrium were not the Marshallian ones. The coincidence between a monopoly firm's option-value thresholds and a competitive firm's free entry threshold was first noted by Leahy (1992). Our exposition mostly follows Dixit (1993b).

Dumas (1992) presented an early model of general equilibrium where each firm made costly switching decisions under Brownian motion uncertainty; this was in the context of international relocation of productive capital, and the aim was to characterize the equilibrium dynamics of real exchange rates.

Caballero (1991) attempted to study industry equilibrium by the shortcut of modelling just one firm's decision, parametrically varying the price elasticity of its demand, and interpreting perfect competition as the limit when the elasticity goes to infinity. He argued that the traditional present value criterion would be restored in this limit. However, one must recognize that the level of price about which to make the demand more elastic is itself endogenous,

and moreover, it acts as a ceiling or reflecting barrier. These effects can be understood only by conducting a proper industry-level analysis. See Pindyck (1993a) for a more specific and detailed discussion of this point. Caballero and Pindyck (1992) later joined forces to produce the model of industry equilibrium with combined firm-specific and aggregate uncertainty that forms the basis of our treatment above.